

## Repondérer une enquête pour corriger la non-réponse : une comparaison de méthodes sur données réelles

Thomas Deroyon, Insee, DMS

La baisse générale des taux de réponse et la complexification de la collecte des enquêtes en face à face accroissent l'importance des redressements, et notamment de la correction de la non-réponse. Celle-ci, dans le cas d'une enquête classique dont l'échantillon est sélectionné suivant un plan de sondage aléatoire, a pour but de supprimer ou limiter le biais d'estimation qu'introduit le fait que l'ensemble des personnes échantillonnées ne fournissent pas de réponses à l'enquête. Pour ce faire, l'enjeu est de mobiliser, de la manière la plus pertinente possible, le maximum d'information auxiliaire. Quand la non-réponse est traitée par repondération, la correction implique d'estimer le plus correctement possible la probabilité de réponse sous-jacente de chaque unité de l'échantillon en s'appuyant sur cette information auxiliaire. Les techniques d'apprentissage statistique dans un cadre supervisé semblent, pour cet objectif, des méthodes intéressantes à étudier.

Cette présentation décrit les travaux réalisés dans le prolongement des expérimentations présentées lors du précédent colloque francophone sur les sondages de 2016, à Gatineau. Elle s'appuie également sur les travaux réalisés par B. Gelein et présentés lors des journées de méthodologie statistique de 2018. L'objectif de ces travaux est d'appliquer à des données réelles d'enquêtes auprès des entreprises et des ménages des méthodes courantes d'apprentissage statistique (forêts aléatoires, boosting,...) et d'évaluer, sur ces données, les gains qu'elles peuvent apporter pour la correction de la non-réponse, en s'appuyant sur des critères d'estimation de la qualité de prédiction du comportement de réponse et de mesure du biais d'estimation.